Some Back-of-the-envelope Calculations for Surveillance Testing

Anonymouse

11/09/2020

About This Document

This document outlines some simple strategies to assess the prevalence of coronavirus when there is insufficient capacity to test everyone in the population of interest. The problem of trying to determine whether a given individual has coronavirus is fundamentally different from the problem of estimating the prevalence of coronavirus in a population of interest. The high specificity and sensitivity along with rapid turnaround that are crucial for individual medical diagnosis are not absolutely essential for surveillance testing at the population level. The purpose of this document is to suggest that only relatively modest resources may be sufficient to obtain valuable information regarding the spread and prevalence of coronavirus in the Oxford population. We view this issue as separate from and parallel to the the "Early Alert Testing" service, whose primary purpose is medical. The ideas we discuss are neither novel nor original, and this document does not propose a detailed plan for surveillance testing. Our goals are merely to explain the intuition behind randomized testing and pooled testing for a non-specialist audience, and to give a rough approximation of the number of tests required to provide useful information. To keep the discussion as simple as possible, we begin by abstracting away the problem of false negatives and false positives. The final section discusses how these affect our analysis.

The Bottom Line

Say something like x number of tests per week could reasonably work if prevalence is such-and-such (with and without pooling)

Random Testing

As a rough approximation, suppose that there are 20,000 undergraduate and taught masters students in residence at Oxford during MT. An unknown number C of them have the coronavirus. With a perfect test, we could learn C exactly by conducting 20,000 tests. But

what if we only have the capacity to carry out n tests, where n is much smaller than 20,000? In this case we cannot learn C exactly, but by employing *random testing* we can produce an estimate that is sufficiently accurate to guide our mitigation policies.

The most basic form of random testing relies on a *simple random sample.* To use this procedure, we place all 20,000 student id numbers into a hat, mix them up thoroughly, and then draw out n at random. These n students are sent for a coronavirus test. We then use the number of students in our *sample* who have coronavirus to estimate the number of students in the *population* who have coronavirus. For example, if one student in a sample of 2000 tests positive, then we would estimate that there are there are 10 cases in the University as a whole. Expressed as a formula, if P is the number of students in our sample who test positive, then our estimate E of the total number of cases in the university equals P/n multiplied by 20,000. In the example this equals 10.

Because students are chosen at random, the estimated number of cases E may not equal the true number of cases C. Purely by chance, we might draw a sample of students in which no one is infected. In this case E would equal zero: an underestimate of C. We could also be unlucky in the opposite direction and, purely by chance, draw a sample that contains *everyone* at Oxford who is infected. In this case E would be an overestimate of C. While an estimate based on a random sample cannot provide us with perfectly accurate information about the true number of cases, with an appropriate sample size it can tell us with high confidence, say, whether the number of cases is larger than 50 or smaller than 20. Furthermore, it can be used to determine whether the prevalence is increasing and, if so, how quickly.

A crucial advantage of random testing over the Early Alert Testing system, which tests only symptomatic individuals, is that random testing it will also detect both pre-symptomatic and asymptomatic cases. The SAGE report of 3rd September refers to a study suggesting that, among undergraduate-aged individuals, asymptomatic cases may constitute up to 80% of total cases. For this reason, even if their turn-around time were somewhat slower, randomized tests could still provide more timely information about the spread of infection.

The key question in designing a random testing protocol is how large a sample size to use. If there are C total cases out of 20,000 students in the university, then on average we would need to carry out approximately $20{,}000/C$ tests to detect a positive case.[1] The more prevalent the infection, the fewer tests we need before we detect a positive.

Table 1: Expe
case in a popu

#Cases	Prevalence (%)
10	0.05
20	0.10
50	0.25
100	0.50
200	1.00

---

[1] The exact value is $20{,}001/(C+1)$.

#Cases	Prevalence (%)
For examp	le, suppose that we can carry out 200 random tests per day. Then within five days we can

As another example, suppose that we initially believed that $C = 20$ and suspected that cases might be doubling every week. With 200 tests per day, we would have a roughly even chance of detecting this, with high confidence, in one week, and be almost certain to detect it within two. (FOOTNOTE)

Approximately 200 test per day would already be useful for answering basic population-level questions concerning prevalence and spread. This would amount to around half as many tests over the entire term as a *single* round of universal testing.

A simple random sampling procedure is the easiest to explain and to use for back-of-the-envelope calculations, but there are several ways that it could be refined to increase the accuracy of our estimates for the same number of daily tests. It is likely that coronavirus cases will tend to "cluster" within colleges, since the infection is most easily spread through frequent, close contact. If this is so, then a sampling procedure that allocates a *fixed* number of tests to each college (proportional to its size) and draws a simple random sample *within* each college will be more efficient than a simple random sample taken at the university level. This is called a *stratified* random sample. Another possibility is to employ *adaptive sampling*, in which additional tests are allocated to colleges in which positive results are detected in a first batch of tests. While the analysis of such a design is more complicated mathematically, it makes more efficient use of a limited number of tests if cases tend to cluster within colleges.

Pooled Testing

Pooled testing is a procedure that can be helpful in situations where laboratory capacity, rather than the ability to collect test samples, is the limiting factor in expanding testing. In this procedure, a single laboratory test is carried out on the *combined* samples of a group of individuals. For a sufficiently sensitive test, a negative result for the pooled sample indicates that no one in the group is infected. Only if the pooled sample tests positive is it necessary to test the individual samples. When the infection is sufficiently rare, pooled testing is highly efficient because virtually all pooled samples will test negative. (Footnote: if you're mainly interested in isolating rather than prevalence, don't even need to re-test! So then just divide by batch size.) For example, if $C = 20$ then pooling tests by household into groups of six results requires approximately 2.3214×10^4

```
p1 <- 20/20000
p2 <- 100/20000
k <- 6
c(1 / k + 1 - (1 - p1)^k, 1 / k + 1 - (1 - p2)^k)
```

```
## [1] 0.1726517 0.1962942
```

What about false positives and negatives?

```
## [1] 0.8885903
```

```
## [1] 0.4724352
```

```
## [1] 0.8721565
```

```
## [1] 0.9935297
```

This latter point is best understood using an example. Suppose that there are 40 coronavirus cases among 20,000 students. This amounts to a prevalence of 0.2%, which would be equivalent to two doublings of the ONS estimate of 0.05% from late August. Accounting for the fact that cases are currently more common among younger age groups, this may be a reasonable approximation to the state of play near the beginning of MT. The following figure shows the different values that the estimated number of cases $\widehat{C}$ could take on, along with their probabilities. In statistical parlance, this is called the "sampling distribution" of $\widehat{C}$. We can calculate these probabilities *exactly* because the randomness in $\widehat{C}$ is entirely under our control: it comes from the fact that students are randomly chosen to be tested.
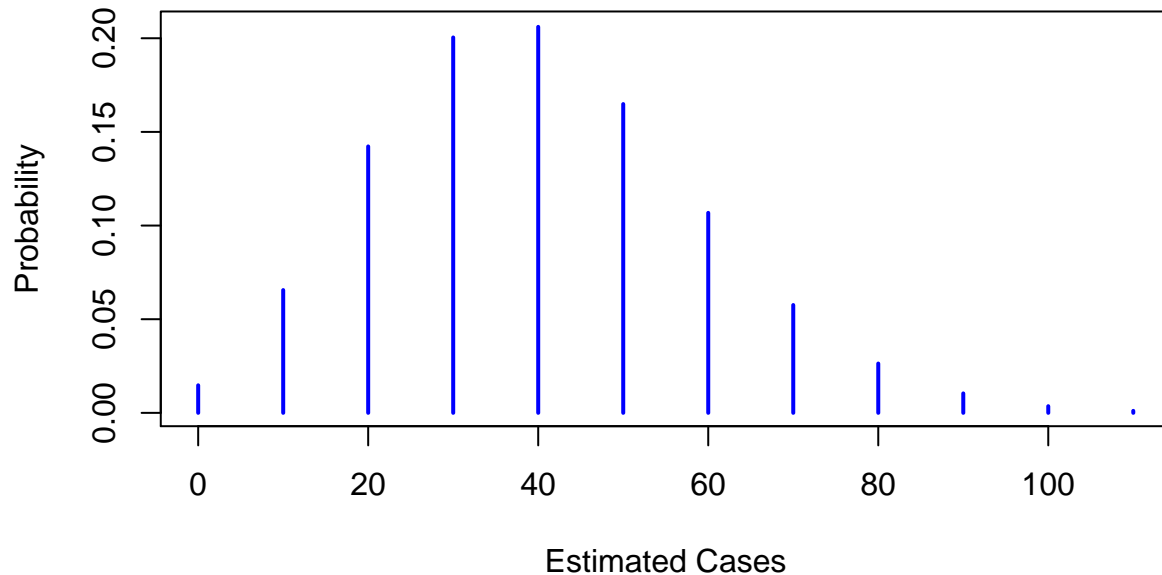

True Cases: 40, Sample Size: 1000
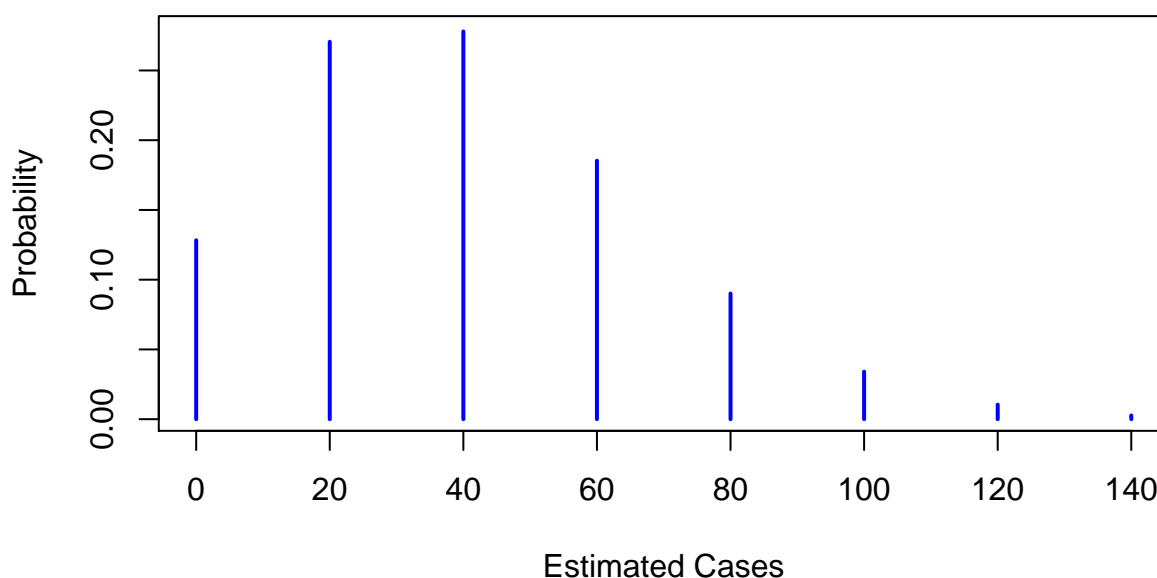
```
d_C_hat <- function(x, n, C, N = 20000) {
  dhyper(n * x / N, C, N - C, n)
}
```

From the figure, we see that $\widehat{C}$ there is a 28% chance that our estimate $\widehat{C}$ will equal 40, the true number of cases. Moreover, there is a 73% chance that it will be between 20 and 60. There is, of course a chance that our sample will not contain any positives: $\widehat{C}$ equals zero with 13% probability.

For a given number of true cases in the population, increasing the sample size makes it more

likely that $\widehat{C}$ will be close to C. For $n = 2000$ the sampling distribution of $\widehat{C}$ becomes more "bell-shaped." With this increased sample size, there is now a 82% chance that it will be between 20 and 60. The chance that we will fail to detect any cases falls to 1%.

True Cases: 40, Sample Size: 2000

For a fixed sample size, the probability of failing to detect any cases also falls as the number of true cases rises. Suppose, for example, that the number of cases in the population were to double from 40 to 80. Then, with a sample size of 1000, the probability that $\widehat{C} = 0$ would fall from 13% to 2%. As a rough guide, a sample size between one and two thousand tests per week should be sufficiently informative to detect cases when the prevalence is 0.2% or above.

True Cases: 80, Sample Size: 1000